SWAT 125: Comparison of trial-collected and routinely-collected death data

Objective of this SWAT
To compare death data collected in trials with the associated routinely-collected health data (RCHD).

Study area: Outcomes, Follow-up, Data Quality
Sample type: Participants
Estimated funding level needed: Low

Background
A promising approach to improving the speed and efficiency, and to reducing the cost, of clinical trials is to make good use of routinely-collected health data (RCHD) in place of sites completing the trial-specific case report forms.[1]

There are many issues to address in order to determine where this is possible, including administrative considerations (e.g. being able to access, store and use RCHD as trial data, and being able to access RCHD in a sufficiently-timely and cost-proportionate manner) and utility considerations (e.g. whether RCHD is source data, and whether the RCHD is as at least as good as trial data).

Trials which already access and collect the same variables through both trial-specific and RCHD routes will allow us to make these comparisons. Some comparisons have already been done,[2,3] but no single trial will provide reassurance or refutation that RCHD should or should not be used for any particular disease setting, outcome measure or time period. Therefore, a patchwork series of SWATs comparing trial data and RCHD would provide a foundation for guidance on where it is possible to use particular RCHD in particular trials. This can be done within ongoing trials without the need to reveal unblinded, accumulating interim data and without a delay in reporting until the end of a trial.

The eligibility for the host trials would be 1) randomised phase III trial; 2) accessed routine-collected health data from national sources (eg NHS Digital or Public Health England, national registries, or audits); 3) death data collected through both trial-specific data collection and RCHD with overlapping or identical fields (fact of death, date of death, cause of death, date death information available); 4) minimum of 20 deaths (to protect patient confidentiality); 5) separate, uncontaminated data sets are available (ie comparison is before merging and before any attempts to reconcile any discrepancies between datasets) (eg where trial data is stored before the knowledge of a death in RCHD prompts trial site to investigate); and 6) trial-specific dataset and RCHD data are appropriately close in time (and ideally from the same data freeze day).

Interventions and comparators

Index Type: Participant outcome data

Method for allocating to intervention or comparator

Outcome measures
Primary: Completeness, agreement and timeliness of the data from the two sources.
Secondary:

Analysis plans
For each trial dataset RCHD dataset pair, the following will be analysed and presented:

1) Linkage: Summarise patients who are not linked to RCHD and therefore would need to be excluded from the trial analysis.

2) Agreement of death: Summarise the number of deaths, current survivors and, if necessary, patients with indeterminate status in the trial dataset and RCHD dataset; assess numbers and percentages of deaths recorded in both datasets combined and in each dataset separately; consider when previously unreported ("missing“) deaths appears in a data source.

3) Agreement of detail: Where a death is reported in both sources, compare the data value in each source in respect to date and cause of death and the potential impact of this on the trial’s analysis

4) Timeliness: Compare the interval between the death data being available in the two sources (using the date of download as an approximation, if necessary).

5) Maturity of follow-up: Calculate median follow-up in the two data sources, using the reverse Kaplan-Meier method.

The results of this SWAT will show, for each host trial, information about how fact of death, date of death, cause of death, and date death information are collected (e.g. frequency and type of contact with patient (e.g. hospital appointment, patient opt in reply). There will also be a conclusion on whether the trial could have used RCHD for the fact, date and/or cause of death, which would be based on the results from the analysis and expressed relative to the data collection frequency used in the host trial.

Possible problems in implementing this SWAT
If there are many time points in the host trial at which trial-specific and RCHD datasets are both available, then this SWAT might need to be limited to the first pair of datasets or the most recent non-cross-contaminated datasets. Issues in implementation by other trial units can be discussed with mrcctu.trialconduct@ucl.ac.uk

References
1) Love SB, Lensen S, Macnair A, et al. Are routinely-collected health data replacing case report forms? A systematic review. Preprint at medRxiv 2020. doi: 10.1101/2020.04.08.20033373.
2) Love SB, Lensen S, Kilanowski A, et al. Routine electronic health records used as participant data in UK randomised trials: the BOSS trial as a case study. Trials 2019; 20(Suppl 1): P-42.
3) Thomas DS, Gentry-Maharaj A, Ryan A, et al. Colorectal cancer ascertainment through cancer registries, hospital episode statistics, and self-reporting compared to confirmation by clinician: A cohort study nested within the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). Cancer Epidemiol 2019; 58: 167-74. doi: 10.1016/j.canep.2018.11.011.

Publications or presentations of this SWAT design
Love SB, Lensen S, Kilanowski A, et al. Routine electronic health records used as participant data in UK randomised trials: the BOSS trial as a case study. Trials 2019; 20(Suppl 1): P-42.

Examples of the implementation of this SWAT

People to show as the source of this idea: Matthew R Sydes, Sharon B Love, Macey L Murray, Sarah Lensen, Archie Macnair, James Carpenter.
Contact email address: mrcctu.trialconduct@ucl.ac.uk
Date of idea: 21/APR/2020
Revisions made by:
Date of revisions: